5К

УДК 004.415.24

Н.В. Кошкина
Институт кибернетики имени В.М. Глушкова НАН Украины, г. Киев, Украина
K_n_v@ukr.net

Внедрение ЦВЗ в аудиосигналы на основе пакетной вейвлет-декомпозиции и частотного маскирования

Рассмотрен спектральный поход к построению систем с цифровыми водяными знаками для аудиосигналов. Эффективность предложенной слепой схемы создания ЦВЗ основывается на использовании особенностей восприятия звука системой человеческого слуха и анализе влияния на спектральные составляющие сигнала типичных операций обработки.

Введение

Цифровой водяной знак (ЦВЗ) представляет собой некоторую малообъемную дополнительную информацию, незаметно внедряемую в сигнал и содержащую идентификатор владельца авторского права, или данной копии цифрового сигнала, или устройства записи, или, например, коды контроля копирования. Согласно требованиям Международной федерации звукозаписывающей индустрии (IFPI) методы внедрения ЦВЗ в аудиосигналы при определенной полезной нагрузке данных (иногда таковой называют пропускную способность образованного стеганоканала больше, чем 20 бит в секунду) должны быть стойкими к типичным операциям обработки сигналов и атакам, таким как масштабирование во временной области (в пределах 10%), аддитивный и мультипликативный шумы, сжатие MPEG-1 Layer 3, передискретизация, перекантование, цифро-аналоговое и аналого-цифровое преобразование. При этом величина вносимых внедрением искажений ограничивается соотношением сигнал-шум SNR>20dB.

На сегодняшний день предложено достаточно большое количество различных алгоритмов создания ЦВЗ для аудиосигналов. Все их можно разделить на 2 класса: алгоритмы с внедрением ЦВЗ во временной или в частотной области сигнала. Характер искажений, вносимых в аудиосигнал операциями обработки, и инвариант к ним, как правило, легче определить в частотной области. Это привело к активной разработке различных частотных методов маркировки аудиосигналов, использующих дискретное преобразование Фурье [1], [2], субполосное кодирование и модифицированное дискретное косинусное преобразование [3], вейвлет-преобразование [4-6].

В работе [7] мы проанализировали влияние на спектр аудиосигнала кодеков сжатия с потерями. Анализ спектрограмм набора аудиосигналов, полученного из исходного сигнала последовательным сжатием с разными битрейтами[1] – от максимально до минимально возможного для различных стандартов (MPEG-1 Layer 3, MPEG-2/4, Ogg Vorbis, WMA), показал, что области существенных изменений в спектре будут расширяться в направлении от высоких частот к низким. Таким образом, инвариант к сжатию разными алгоритмами целесообразно искать в низких частотах аудиосигналов.

[1] Битрейт – количество битов информации на единицу времени. Например, в стандарте MPEG-1 Layer 3 предусматривается сжатие с битрейтом от 32 до 320 кбит/с.

В работах [8], [9] была представлена новая схема маркировки аудиосигналов, стойкая к сжатию с потерями, передискретизации, низкочастотной фильтрации, аддитивному и мультипликативному шуму и удовлетворяющая при этом условие SNR>20dB. Схема основана на вейвлет-декомпозиции сигнала и внедрении ЦВЗ модификацией определенных амплитуд низкочастотной субполосы. Недостатком, ограничивающим применение этой схемы, является неслепой метод декодирования ЦВЗ, использованный в ней. То есть для извлечения ЦВЗ необходимо наличие некоторой дополнительной ключевой информации, которая формируется алгоритмом внедрения. Несмотря на возможность весьма существенной минимизации длины ключевых данных при функционировании стеганосистемы, сохраняется необходимость хранения и использования ключей, привязанных не к пользователю, а к маркированному сигналу. В данной работе будет очерчена альтернативная схема маркировки, декодирование ЦВЗ в которой является слепым.

Неощутимость, стойкость и вместимость при построении систем с ЦВЗ

Разработка эффективных систем с ЦВЗ упирается в компромисс между тремя их основными характеристиками – неощутимостью (imperceptibility), стойкостью (robustness) и пропускной способностью или вместимостью (capacity). Неощутимости ЦВЗ можно добиться либо внося минимальные модификации в отсчеты сигнала или частотные коэффициенты, либо эксплуатируя эффекты маскировки одних звуков другими. Причем использование маскировки более перспективно в плане стойкости, поскольку небольшие модификации в большинстве случаев искажаются операциями обработки.

Маскировка одних звуков другими проявляется как во временной области, так и в частотной. На эффекте временной маскировки построен, например, метод внедрения ЦВЗ эхо-кодированием [10]. Традиционно для того, чтобы использовать эффект частотной маскировки и рассчитать пороги маскирования, модификации ниже которых неощутимы человеком, частотные коэффициенты вычисляют при помощи быстрого преобразования Фурье (БПФ). В частности этот подход эксплуатируется в психоакустической модели стандарта сжатия с потерями MPEG-1 Layer 3 [11]. Вместе с тем в последние годы стали появляться новые психоакустические модели, использующие для анализа аудиосигнала пакетную вейвлет-декомпозицию [12], [13]. По сравнению с традиционными моделями они способны обеспечить мультиразрешающий анализ сигнала и более точно аппроксимировать разбиение его частотного диапазона согласно критическим полосам слуха. Такие новые психоакустические модели востребованы как для разработки алгоритмов сжатия с потерями, так и для алгоритмов маркировки аудиосигналов.

Возможно также объединение сжатия и маркировки в один процесс. В этом случае биты ЦВЗ внедряются в сигнал методом модуляции индекса квантования [14], в котором использовано два квантователя. Один квантователь применяется при внедрении нуля, второй – при внедрении единицы. Шаг квантования вычисляется на основе психоакустической модели, контролирующей неслышимость шума квантования.

Метод модуляции индекса квантования является обобщением широко известного метода наименьшего значащего бита. Вместе с тем в отличие от первого он позволяет выполнить адаптивный подбор шага квантования и таким образом добиться повышения стойкости. С другой стороны, метод модуляции индекса квантования шумный на низкоамплитудных участках. Аналитически метод исследовался в работах [1], [15], где была показана его уязвимость к дальнейшим атакам.

При соблюдении требования неощутимости повысить стойкость можно понижая вместимость, т.е. внедрять биты ЦВЗ не во весь сигнал, а только в те его участки, которые после возможных операций обработки сохранятся точно или приблизительно в таком же виде, как в исходном аудио. Учитывая эффект частотной маскировки, который наиболее сильно проявляется внутри критических полос слуха, такими существенными участками целесообразно выбрать участки вокруг низкочастотных тональных маскирующих компонентов, определяемых психоакустической моделью. Бит ЦВЗ будет закодирован целенаправленным формированием разности амплитуд левой и правой соседних к выбранному тону частотных составляющих.

Неощутимость вносимых изменений базируется на наличии скрытого порога маскирования, инициированного данным тональным компонентом. Стойкость – на перцепционной важности тональных компонент в сигнале. Перед применением процедуры внедрения аудиосигнал разбивается на сегменты по несколько секунд звучания. Вместимость стеганоканала регулируется длиной сегмента и числом тех субполос в нем, в которые будут внедряться биты ЦВЗ.

Аппроксимация критических полос слуха деревом пакетной вейвлет-декомпозиции сигнала

В психоакустике принято выделять 25 критических полос слуха (табл. 1), внутри которых происходит интегрирование поступившей звуковой информации и эффект частотной маскировки наиболее выражен.

Таблица 1 – Разбиение слышимого диапазона частот на критические полосы

№ полосы	Центральная частота, Гц	Нижняя и верхняя частоты полосы, Гц	Ширина полосы, Гц
1	50	20 – 100	100
2	150	100 – 200	100
3	250	200 – 300	100
4	350	300 – 400	100
5	450	400 – 510	110
6	570	510 – 630	120
7	700	630 – 770	140
8	840	770 – 920	150
9	1000	920 – 1080	160
10	1170	1080 – 1270	190
11	1370	1270 – 1480	210
12	1600	1480 – 1720	240
13	1850	1720 – 2000	280
14	2150	2000 – 2320	320
15	2500	2320 – 2700	380
16	2900	2700 – 3150	450
17	3400	3150 – 3700	550
18	4000	3700 – 4400	700
19	4800	4400 – 5300	900
20	5800	5300 – 6400	1100
21	7000	6400 – 7700	1300
22	8500	7700 – 9500	1800
23	10500	9500 – 12000	2500
24	13500	12000 – 15500	3500
25	19000	155000 – 22500	7000

Хорошую аппроксимацию критических полос слуха можно получить построением согласованного с приведенными данными дерева пакетной вейвлет-декомпозиции сигнала. Каждому ортогональному вейвлету соответствуют его низкочастотный *g* и высокочастотный *h* фильтры. Применение к аудиосигналу данных фильтров с децимацией результата в соотношении два к одному разбивает сигнал на две субполосы – низкочастотную и высокочастотную. Каждая из субполос этими же фильтрами может быть разбита еще на две части. Декомпозицию можно продолжать вплоть до момента, когда субполосы разложения будут содержать по одному отсчету. Такой подход использован в алгоритмах быстрого вейвлет-преобразования и является сутью многоуровневой вейвлет-декомпозиции сигнала на частотные субполосы.

Процедура расчета параметров психоакустической модели в области вейвлет-коэффициентов описана, например, в работах [12], [13]. В статье [12] для аппроксимации критических частотных полос предложено следующее дерево пакетной вейвлет-декомпозиции (рис. 1).

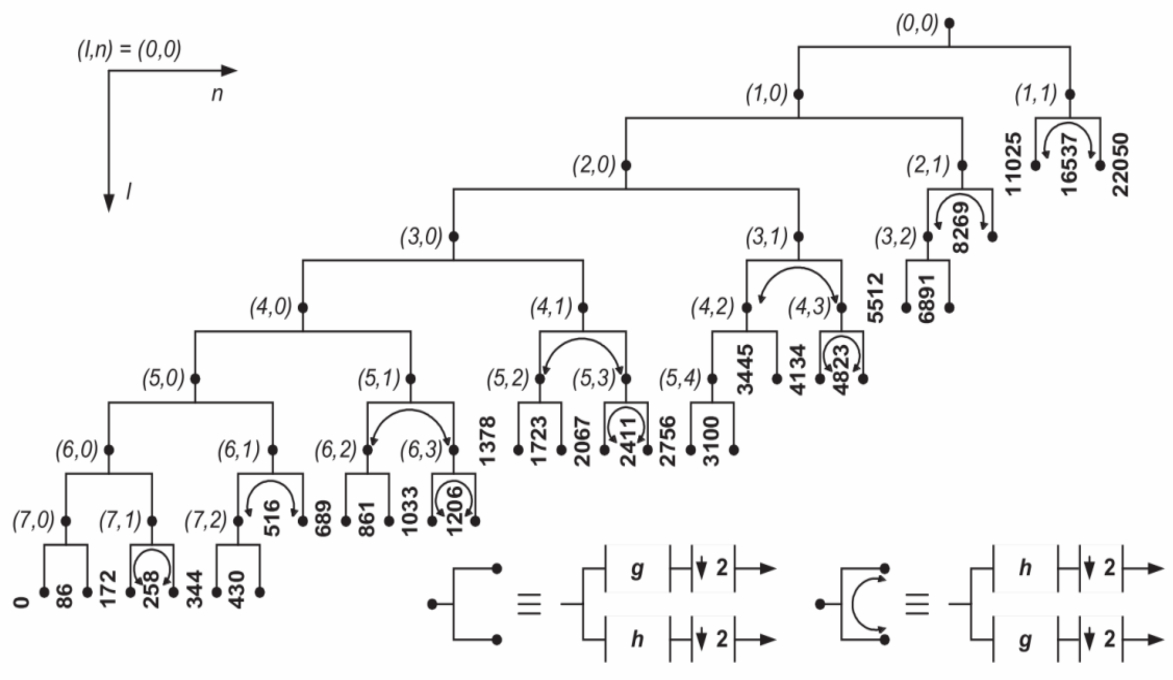


Рисунок 1 – Дерево пакетной вейвлет-декомпозиции сигнала, аппроксимирующее критические полосы слуха

Здесь (l,n) – нумерация узлов дерева декомпозиции, а вертикально проставленные цифры в конечных узлах – частоты в Гц, соответствующие каждой из субполос разложения, т.е. первая субполоса будет содержать отсчеты с частотами от 0 до 86 Гц, вторая – от 86 до 172 Гц и т.д.

Таким образом, для аппроксимации критических полос в аудиосигнале, оцифрованном с частотой дискретизации 44 кГц, используется 8 уровней пакетного вейвлет-разложения. В качестве базисного вейвлета, как правило, выбирают вейвлеты Добеши. При этом чем выше порядок выбранного вейвлета, тем более тонкую структуру сигнала он позволяет анализировать. Так, например, в работе [13] при построении психоакустической модели используется вейвлет Добеши 8 порядка.

В целях обеспечения стойкости к атакам внедрение битов ЦВЗ должно выполняться только в первые 10 – 20 частотных субполос.

Использование частотного маскирования при внедрении битов ЦВЗ

Рассмотрим аудиосигнал в частотной области. На рис. 2 проиллюстрирован эффект маскирования соседних частотных компонент сильным тоном. Согласно существующим исследованиям [2], если $F(f_k)$ – сильный тональный компонент, то в пределах частотной полосы с шириной, меньшей половины критической, перераспределение энергии вида

$$P_{TM}(f_k) = 10\log_{10}\left(\left|F(f_k)\right|^2 + \left(\left|F(f_{k-1})\right|^2 + \Delta e\right) + \left(\left|F(f_{k+1})\right|^2 - \Delta e\right)\right)$$

остается неслышимым.

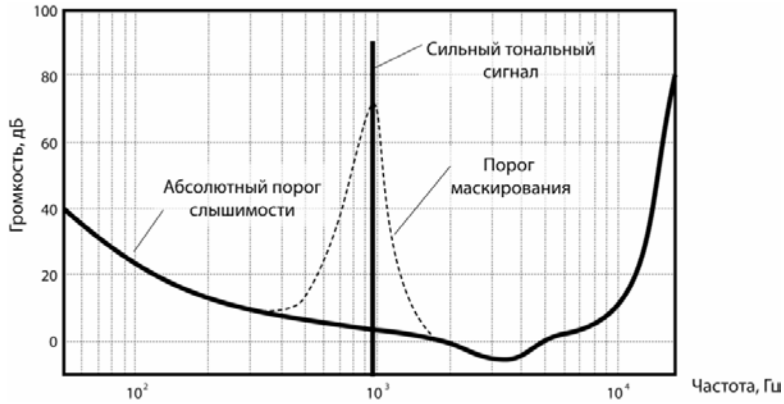


Рисунок 2 – Абсолютный порог слышимости и порог маскирования, возникновение которого инициирует чистый тон с частотой 1 кГц

С другой стороны, результаты экспериментов по оценке изменения энергии тона $E_t = \left|F(f_k)\right|^2$, энергии левой и правой, соседних с тоном частот $E_l = \left|F(f_{k-1})\right|^2$ и $E_r = \left|F(f_{k+1})\right|^2$, а также отношения энергий E_l / E_t и E_r / E_t под влиянием типичных операций обработки сигнала выявили их достаточную стабильность.

Это позволяет внедрить ЦВЗ, манипулируя разностью между левым и правым соседними с тональным частотными компонентами.

Внедрение битов ЦВЗ

Тройку частотных отсчетов $F(f_{k-1})$, $F(f_k)$, $F(f_{k+1})$, где $F(f_k)$ – сильный тональный компонент, будем называть тональным маскером.

Для внедрения нулевого бита ЦВЗ нужно обеспечить соотношение:

$$R(E_l, E_r) \le L_1, \tag{1}$$

а единичного:

$$R(E_l, E_r) \ge L_2, \tag{2}$$

где $R(E_l, E_r) = \left|E_l - Er\right|/2$, $L_1 = (E_t - E_m)\cdot\alpha$, $L_2 = (E_t - E_m)\cdot\beta$, $E_m = (E_l + E_r)/2$. α и β – константы, регулирующие взаимосвязь между неощутимостью и стойкостью внедрения. $\alpha < \beta$, и для стойкости разница между ними должна быть как можно больше. Вместе с тем слишком большая разница повлечет за собой слышимые искажения в сигнале.

Если соотношения (1) и (2) изначально справедливы для некоторого тонального маскера исходного сигнала, в процессе маркировки не вносится никаких дополнительных изменений. Но если при внедрении нулевого бита в исходном сигнале имеют $R(E_l, E_r) > L_1$, то выполнения соотношения (1) добиваются следующим образом:

$$E_l = E_m + L_1 \wedge E_r = E_m - L_1, \text{ при } E_l \geq E_r;$$
$$E_l = E_m - L_1 \wedge E_r = E_m + L_1, \text{ при } E_l < E_r.$$

Аналогично, если при внедрении единичного бита имеют $R(E_l, E_r) < L_2$, то соотношение (2) обеспечивается так:

$$E_l = E_m + L_2 \wedge E_r = E_m - L_2, \text{ при } E_l \geq E_r;$$
$$E_l = E_m - L_2 \wedge E_r = E_m + L_2, \text{ при } E_l < E_r.$$

При таком подходе важно найти устойчивые тональные компоненты, которые будут однозначно идентифицироваться до и после искажений, внесенных обработкой аудиосигнала. В случае, когда в субполосе имеется несколько сильных тонов, близких по значению, для повышения надежности идентификации тона-носителя бита ЦВЗ допускается незначительное увеличение выбранного тона-носителя и/или незначительное уменьшение близких тонов.

Извлечение битов ЦВЗ

Пусть W – бит ЦВЗ, носителем которого является тон с энергией E_t. Тогда его извлечение выполняется согласно следующим правилам:

$$W = 0, \text{ при } R(E_l, E_r) \leq L_1;$$
$$W = 1, \text{ при } R(E_l, E_r) \geq L_2.$$

В случае, если $L_1 \leq R(E_l, E_r) \leq L_2$:

$$W = 0, \text{ при } ((R(E_l, E_r) - L_1) - L_3) / L_3 > 1/2;$$
$$W = 1, \text{ при } ((R(E_l, E_r) - L_1) - L_3) / L_3 < 1/2.$$

Схематически процесс внедрения-извлечения бита ЦВЗ можно проиллюстрировать так (рис. 3).

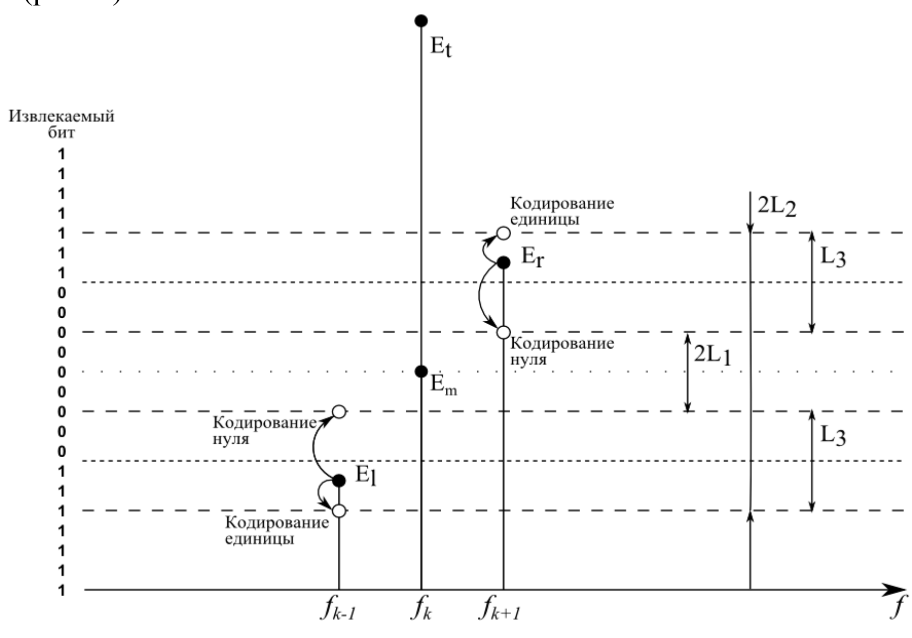


Рисунок 3 – Схема внедрения-извлечения одного бита ЦВЗ

Заключение

Проблематика защиты информации актуальна всегда. ЦВЗ являются эффективной технологией защиты авторских прав на цифровые объекты, контроля копирования и распространения данных. В данной статье изложено новое, перспективное направление исследований в области создания ЦВЗ для аудиосигналов. Оно перекликается с направлениями последних разработок в области сжатия звуковых и речевых сигналов, где так же, как и в стеганографии, актуально построение качественных моделей слухового восприятия. Сжатие с потерями направлено на устранение психоакустической и статистической избыточности сигнала. Задача создания ЦВЗ в целом тоньше, «ювелирнее» задачи сжатия с потерями. Внедрение ЦВЗ, как и применение алгоритма сжатия, не должно ухудшать качество восприятия аудиосигнала. Вместе с тем в отличие от психоакустического сжатия, потери данных (содержимого ЦВЗ), в системе маркировки аудио крайне нежелательны, в том числе после сжатия маркированного сигнала и других его преобразований. Таким образом, для целей стеганографии очень актуально усовершенствование существующих психоакустических моделей и средств их представления. И один из путей такого усовершенствования – использование вейвлет-пакетов.

Психоакустическая модель в данном случае строится для выделения тональных маскеров, целенаправленным формированием которых кодируется битовая последовательность ЦВЗ.

Литература

1. Quantization-Based Digital Audio Watermarking in Discrete Fourier Transform Domain / S. Yang, W. Tan, Y. Chen, W.Ma // Journal of Multimedia. – 2010. – V. 5, № 2. – P. 151-158.
2. Lee H.S. and Lee W.S. Audio watermarking through modification of tonal maskers / H.S. Lee, W.S. Lee // ETRI Journal. – 2005. – V. 27, № 5. – P. 608-616.
3. Saito S., Furukawa T., Konishi K. A digital watermarking for audio data using band division based on QMF bank / S. Saito, T. Furukawa, K. Konishi // IEEE International Conference on Acoustics, Speech, and Signal Processing. – 2002. – V. 4. – P. 3473-3476.
4. Fu Y., Ma Z., Song G. A robust audio watermarking algorithm based on wavelet transform // Journal of Information and Computational Science 2. – 2005. –№ 1. – P. 7-11.
5. S. Lee, S.-K.Lee, Y.-H. Seo, C. Yoo. Capturing-Resistant Audiowatermarking based on Discrete Wavelet Transform //In Proceedings of ICME. – 2006. – P. 1105-1108.
6. Al-Haj A., Mohammad A. Digital Audio Watermarking Based on the Discrete Wavelets Transform and Singular Value Decomposition / A. Al-Haj, A. Mohammad // European Journal of Scientific Research ISSN 1450-216X. – 2010. – V. 39, № 1. – P. 6-21.
7. Кошкина Н.В. Определение инварианта к сжатию с потерями для аудиосигналов / Н.В. Кошкина // Управляющие системы и машины (УСиМ). – 2010. – № 3. – С. 86-93.
8. Кошкина Н.В. Побудова стеганосистем з ЦВЗ для аудіо сигналів / Н.В. Кошкина // Поступ в науку : матеріали конференції ПНМК-2010. – 2010. – Т. 1, № 6. – С. 117-120.
9. Кошкина Н.В. Метод встраивания ЦВЗ в аудиосигналы на основе вейвлет- и Фурье преобразований / Н.В. Кошкина // Проблемы управления и информатики. – 2010. – № 6 (в печати).
10. Techniques for Data Hiding / W. Bender, D. Gruhl, N. Morimoto, A. Lu // IBM Systems J. – 1996. – V. 35. – P. 313-336.
11. Pan D. A Tutorial on MPEG/Audio Compression / D. Pan // IEEE Multimedia. –1995. – V. 2, № 2. – P. 60-74.
12. Петровский А.А. Построение психоакустической модели в области вейвлет-коэффициентов для перцептуальной обработки звуковых и речевых сигналов //Речевые технологи. – 2008. – №4. – С. 61 -71.
13. He X. Psychoacoustic Music Analysis Based on the Discrete Wavelet Packet Transform / X. He, M.S. Scordilis // Hindawi Publishing Corporation Research Letters in Signal Processing. – 2008. – Article ID 346767, 5 pages, doi:10.1155/2008/346767
14. Chen B. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding / B. Chen, G.W. Wornell // IEEE Transactions on Information Theory. – 2001. – V. 47 – P. 1423-1443.
15. Xiang S., Huang J. Analysis of D/A and A/D Conversions in Quantization-based Audio Watermarking / S. Xiang, J. Huang // International Journal of Network Security. – 2006. – V. 3, № 3. – P. 230-238.

N.V. Koshkina

Audio Watermarking Based on the Discrete Wavelet Packet Transform and Frequency Masking

In this work the spectral method for building of systems with the digital audio watermarks is considered. The effectiveness of the proposed blind watermarking scheme is based on features of human auditory system and also on the analysis of typical processing operations influences to the spectral components of audio signal.

Статья поступила в редакцию 23.06.2010.